Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER

Journal of Computational and Applied Mathematics 170 (2004) 349–370

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

A gradient-related algorithm with inexact line searches ☆

Zhen-Jun Shi[a,b,*], Jie Shen[c]

[a] *College of Operations Research and Management, Qufu Normal University (Rizhao Campus), Rizhao, Shandong 276826, PR China*

[b] *Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100080, China*

[c] *Department of Computer & Information Science, University of Michigan, Dearborn, MI 48128, USA*

Received 10 March 2003; received in revised form 27 October 2003

Abstract

In this paper, a new gradient-related algorithm for solving large-scale unconstrained optimization problems is proposed. The new algorithm is a kind of line search method. The basic idea is to choose a combination of the current gradient and some previous search directions as a new search direction and to find a step-size by using various inexact line searches. Using more information at the current iterative step may improve the performance of the algorithm. This motivates us to find some new gradient algorithms which may be more effective than standard conjugate gradient methods. Uniformly gradient-related conception is useful and it can be used to analyze global convergence of the new algorithm. The global convergence and linear convergence rate of the new algorithm are investigated under diverse weak conditions. Numerical experiments show that the new algorithm seems to converge more stably and is superior to other similar methods in many situations.

MSC: 90C30; 65K05; 49M37

Keywords: Unconstrained optimization; Gradient-related algorithm; Inexact line search; Convergence

1. Introduction

Consider an unconstrained optimization problem (UP)

$$\min f(x), \quad x \in \mathbb{R}^n, \tag{1}$$

☆ The work is partly supported by Natural Science Foundation of China (grant 10171054), Postdoctoral Foundation of China and Kuan-Cheng Wang Postdoctoral Foundation of CAS (grant 6765700).

* Corresponding author. Tel.: +86-0633-3980323; fax: +86-0633-3980329.

E-mail addresses: zjshi@qrnu.edu.cn (Z.-J. Shi), shen@umich.edu (J. Shen).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2003.10.025

where $f:\mathbb{R}^n \to R^1$ is a continuously differentiable function, $\mathbb{R}^n$ is an n-dimensional Euclidean space and n may be very large in some sense.

These problems often arise not only from many application fields such as economic, social, science, engineering, management fields, etc. [14,23], but also from many theoretical fields because most of optimization problems can be reduced to an unconstrained optimization problem [20,22].

Most of well-known iterative algorithms for solving (UP) take the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where d_k is a search direction and α_k a positive step-size along the search direction. This class of methods is called line search gradient method (e.g., [24,31], etc.). If x_k is the current iterative point, then we denote $\nabla f(x_k)$ by g_k, $f(x_k)$ by f_k and $f(x^*)$ by f^*, respectively.

If we take $d_k = -g_k$, then the corresponding method is called steepest descent method, a simple one in gradient methods. It has a wide application in large-scale optimization fields. However, steepest descent method often yields zigzag phenomena in solving practical problems, which sometimes makes the algorithm converge very slowly, or even fail to converge [9,6,14,20,22,23].

Generally, conjugate gradient method is a useful technique for solving large-scale problems because it avoids the computation and storage of some matrices associated with Hessian of objective functions. The conjugate gradient method has the form

$$d_k = \begin{cases} -g_k & \text{if } k = 1, \\ -g_k + \beta_k d_{k-1} & \text{if } k \geqslant 2, \end{cases} \tag{3}$$

where β_k is a parameter that determines the different conjugate gradient methods [5,7,11,12,15,16]. For example, well-known choices of β_k can be taken as

$$\beta_k^{\mathrm{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{\mathrm{PRP}} = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \quad \beta_k^{\mathrm{HS}} = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{d_{k-1}^{\mathrm{T}} g_{k-1}},$$

which, respectively, correspond to the FR (Flecther–Revees [10]), PRP (Polak–Ribière–Polyak [19,25,26]) and HS (Hestenes–Stiefel [20]) conjugate gradient methods.

Fletcher [9] presented a conjugate descent method (abbreviated as CD) in which β_k was defined by

$$\beta_k^{\mathrm{CD}} = -\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}} g_{k-1}}.$$

Dai and Yuan [6] used the following formula for β_k in (3):

$$\beta_k^{\mathrm{DY}} = \frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})}.$$

Conjugate gradient methods with exact line search have finite convergence when they are used to minimize strictly convex quadratic functions ([8,10,16], etc.). However, if objective function is not quadratic or exact line search is not used, a conjugate gradient method has no finite convergence in many situations. Moreover, many conjugate gradient methods have no global convergence if the objective function is nonquadratic, although this class of methods often has good performance in practical computation. Large literatures on conjugate gradient methods have appeared in recent decades (e.g., [5,7,8,11,15,16,29], etc.).

Similarly, Miele and Cantrell [21] studied memory gradient method for (UP). If x_0 is an initial point and $\delta_0 = g_0$, the algorithm can be stated as follows:

$$x_{k+1} = x_k + \delta_k, \quad \delta_k = -\alpha g_k + \beta \delta_{k-1},$$

where α and β are scalars chosen at each step so as to yield the greatest decrease in the function f. Cantrell [3] showed that the memory gradient method and the Fletcher–Reeves algorithm [10] were identical in the particular case of a quadratic function.

Cragg and Levy [4] proposed a super-memory gradient method as follows:

$$x_{k+1} = x_k + \delta_k, \quad \delta_k = -\alpha g_k + \sum_{i=1}^{k} \beta_i \delta_{i-1},$$

where $\delta_i = x_{i+1} - x_i$, x_0 is an initial point, and $\delta_0 = g_0$.

Wolfe and Viazminsky [32] investigated a super-memory descent method for (UP) in which the main iteration takes the form

$$f\left(x_k - \alpha_k p_k + \sum_{i=1}^{m} \beta_i^{(k)} \delta_{k-i}\right) = \min_{\alpha, \beta_1, \ldots, \beta_m} f\left(x_k - \alpha p_k + \sum_{i=1}^{m} \beta_i \delta_{k-i}\right),$$

where

$$x_{k+1} = x_k + \delta_k, \quad \delta_k = -\alpha_k p_k + \sum_{i=1}^{m} \beta_i^{(k)} \delta_{k-i},$$

m is a fixed positive integer, and $p_k^{\mathrm{T}} g_k \neq 0$.

Both memory and super-memory gradient methods are more efficient than conjugate gradient methods [27,29,30], but the memory or super-memory gradient methods demand much greater amount of computation and storage than conjugate gradients and steepest descent methods. We may combine conjugate gradient methods and super-memory gradient methods to devise some new algorithms for solving large-scale optimization problems.

In this paper, a new gradient-related algorithm for solving large-scale unconstrained optimization problems is proposed. The new algorithm is a kind of line search method. The basic idea is to choose a combination of the current gradient and some previous search directions as a new search direction and to find a step-size by using various inexact line searches. Using more information at the current iterative step may improve the performance of the algorithm. This motivates us to find some new gradient algorithms that may be more effective than standard conjugate gradient methods. Uniformly gradient-related conception is useful and it can be used to analyze global convergence of the new algorithm. The global convergence and convergence rate are investigated under diverse weak conditions. Numerical experiments show that the new algorithm seems to converge more stably and is superior to other similar methods in many situations.

The rest of this paper is organized as follows. Section 2 describes the algorithm and analyzes its simple properties. In Sections 3–5, we prove its global convergence and convergence rate. Numerical experiments and comparisons are given in Section 6.

2. New algorithm

We assume that

(H_1) The objective function f has lower bound on the level set $L_0 = \{x \in \mathbb{R}^n | f(x) \leqslant f(x_0)\}$, where x_0 is an available initial point.

(H_2) The gradient $g(x)$ of $f(x)$ is Lipschitz continuous in an open convex set B which contains L_0, i.e., there exists $L > 0$ such that

$$\|g(x) - g(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in B, \tag{4}$$

where $g(x) = \nabla f(x)$.

(H_3) The gradient $g(x)$ is uniformly continuous in an open convex set B containing L_0.

Obviously, Assumption (H_2) implies (H_3). It will be shown that new algorithm in the paper has weak convergence condition.

As we know, a key to devise an algorithm for unconstrained optimization problems is to choose an available search direction d_k and a suitable step-size α_k at each iteration.

Certainly, if we choose a search direction d_k satisfying

$$g_k^{\mathrm{T}} d_k < 0, \tag{5}$$

then we can devise a descent method [20–24]. Generally, we demand that

$$-g_k^{\mathrm{T}} d_k \geqslant \eta \|g_k\|^2 \tag{6}$$

which is called sufficient descent condition [5,12], where $\eta > 0$. Furthermore, if

$$-g_k^{\mathrm{T}} d_k \geqslant \eta \|g_k\| \cdot \|d_k\| \tag{7}$$

then many descent algorithms have convergence [5,14,16,18,20,22,24]. The above condition is called angle condition. There is also a so-called gradient-related conception.

Definition 2.1 (Bertsekas [2]). Let $\{x_k\}$ be a sequence generated by a gradient method $x_{k+1} = x_k + \alpha_k d_k$. We say that the sequence $\{d_k\}$ is uniformly gradient related to $\{x_k\}$ if for every convergent subsequence $\{x_k\}_K$ for which

$$\lim_{k \in K, k \to \infty} g_k \neq 0$$

there holds

$$0 < \liminf_{k \in K, k \to \infty} |g_k^{\mathrm{T}} d_k|, \quad \limsup_{k \in K, k \to \infty} \|d_k\| < +\infty.$$

In words, $\{d_k\}$ is uniformly gradient related if whenever a subsequence $\{g_k\}_K$ tends to a nonzero vector, the corresponding subsequence of directions d_k is bounded and does not tend to be orthogonal to g_k.

Moreover, we must choose a line search rule to find the step-size along search direction at each iteration ([2,24,31], etc.).

(a) Exact minimization rule: Here α_k is chosen so that

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geqslant 0} f(x_k + \alpha d_k).$$

(b) Limited minimization rule: A fixed number $s > 0$ is selected and α_k is chosen so that

$$f(x_k + \alpha_k d_k) = \min_{\alpha \in [0,s]} f(x_k + \alpha d_k).$$

(c) Armijo rule: α_k is the largest number in $\{s, s/2, s/2^2, \ldots\}$ such that

$$f_k - f(x_k + \alpha d_k) \geqslant -\mu_2 \alpha g_k^{\mathrm{T}} d_k,$$

where $\mu_2 \in (0,1)$, and $s > 0$.

Note: In fact, we can use the generalized Armijo's line search see [24,31]: α_k is the largest number in $\{s, s\beta, s\beta^2, s\beta^3, \ldots\}$ such that

$$f_k - f(x_k + \alpha d_k) \geqslant -\mu_2 \alpha g_k^{\mathrm{T}} d_k,$$

where $\mu_2 \in (0,1)$, $\beta \in (0,1)$ and $s > 0$. How to choose β is important for the implementation of line search methods. If $\beta \in (0,1)$ is too large then line search process may be too slow. While $\beta \in (0,1)$ is too small, line search process may be too fast so as to lose the available stepsize. We should choose a suitable stepsize at each iteration.

(d) Goldstein rule: A fixed scalar $\sigma \in (0, \frac{1}{2})$ is selected, and α_k is chosen to satisfy

$$\mu_1 \leqslant \frac{f(x_k + \alpha_k d_k) - f_k}{\alpha_k g_k^{\mathrm{T}} d_k} \leqslant \mu_2,$$

where $\mu_1 = \sigma$ and $\mu_2 = 1 - \mu_1$. It is possible to show that if f is bounded below there exists an interval of step-sizes α_k for which the relation above is satisfied.

(e) Strong Wolfe search rule: The step-size α_k satisfies simultaneously

$$f(x_k + \alpha_k d_k) - f_k \leqslant \alpha_k \mu_1 g_k^{\mathrm{T}} d_k \tag{8}$$

and

$$|g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k| \leqslant \mu_2 |g_k^{\mathrm{T}} d_k|, \tag{9}$$

where $0 < \mu_1 < \frac{1}{2} < \mu_2 < 1$.

We can see that (a) and (b) are exact line searches. It may be impossible to implement in practice, so that we generally use inexact line search rules in some line search gradient algorithms.

Lemma 2.1 (Berstsekas [1]). *Let $\{x_k\}$ be a sequence generated by a gradient method and assume that $\{d_k\}$ is uniformly gradient related and α_k is chosen by the minimization rule, or the limited minimization rule, or the Armijo rule, or the Goldstein rule, or the strong Wolfe rule. Then every limit point of $\{x_k\}$ is a critical point x^*, i.e., $g(x^*) = 0$.*

We will define the following algorithm and prove that search direction sequence is gradient related to iterative point sequence.

Algorithm (A). $0 < \rho < 1$, $0 < \mu_1 < \frac{1}{2} < \mu_2 < 1$, a fixed integer $m \geqslant 2$, $x_1 \in \mathbb{R}^n$, $k := 1$

Step 1: If $\|g_k\| = 0$ then stop! else goto step 2;

Step 2: $x_{k+1} = x_k + \alpha_k d_k(\beta_{k-m+1}^{(k)}, \dots, \beta_k^{(k)})$, where

$$d_k(\beta_{k-m+1}^{(k)}, \dots, \beta_k^{(k)}) = \begin{cases} -g_k & \text{if } k \leqslant m-1, \\ -\beta_k^{(k)} g_k - \sum_{i=2}^{m} \beta_{k-i+1}^{(k)} d_{k-i+1} & \text{if } k \geqslant m, \end{cases} \tag{10}$$

where $\beta_{k-i+1}^{(k)} \in [s_k^i/2, s_k^i]$, or for $i = 2, \dots, m$,

$$\beta_{k-i+1}^{(k)} = \begin{cases} s_k^i/2 & \text{if } \|g_k\|^2 \geqslant g_k^{\mathrm{T}} d_{k-i+1}, \\ s_k^i & \text{if } \|g_k\|^2 < g_k^{\mathrm{T}} d_{k-i+1} \end{cases}$$

and

$$s_k^1 = 1, \; s_k^i = \frac{\rho}{m-1} \cdot \frac{\|g_k\|^2}{\|g_k\|^2 + |g_k^{\mathrm{T}} d_{k-i+1}|} \; (i = 2, \dots, m), \; \beta_k^{(k)} = 1 - \sum_{i=2}^{m} \beta_{k-i+1}^{(k)},$$

and α_k is chosen by line search rule (a), or (b), or (c), or (d), or (e).

Step 3: $k := k + 1$, goto Step 1.

Note: At each iteration of this algorithm, we use more information to structure search directions and use some parameters to adjust the convergence property. This makes us obtain many efficient and stable algorithms. The numerical experiment will show that the new algorithm can converge more quickly and stably than other similar line search methods such as conjugate gradient methods. As we can see in numerical experiments, how to choose the constants ρ, μ_1, μ_2, and the integer m will influence the implementation efficiency of the new algorithm.

As to the parameters in the algorithm, we seem to choose $\beta_{k-i+1}^{(k)} \in [s_k^i/2, s_k^i]$ $(i = 2, \dots, m)$ such that

$$\min \left\{ g_k^{\mathrm{T}} d_k \left(\beta_{k-m+1}^{(k)}, \dots, 1 - \sum_{i=2}^{m} \beta_{k-i+1}^{(k)} \right) \middle| \beta_{k-i+1}^{(k)} \in [s_k^i/2, s_k^i] \; (i = 2, \dots, m) \right\}$$

so as to make the algorithm converge more quickly. Therefore, we obtain

$$\beta_{k-i+1}^{(k)} = \begin{cases} s_k^i/2 & \text{if } \|g_k\|^2 \geqslant g_k^{\mathrm{T}} d_{k-i+1}, \\ s_k^i & \text{if } \|g_k\|^2 < g_k^{\mathrm{T}} d_{k-i+1}, \end{cases}$$

for $i = 2, \dots, m$.

For simplicity, we denote $d_k(\beta_{k-m+1}^{(k)}, \dots, \beta_k^{(k)})$ by d_k throughout the paper. It is easy to prove that

$$0 < \sum_{i=2}^{m} \beta_{k-i+1}^{(k)} \leqslant \sum_{i=2}^{m} s_k^i \leqslant \rho < 1$$

and

$$1 > \beta_k^{(k)} = 1 - \sum_{i=2}^{m} \beta_{k-i+1}^{(k)} \geqslant 1 - \sum_{i=2}^{m} s_k^i \geqslant 1 - \rho > 0.$$

Lemma 2.2. *For all* $k \geqslant 1$,

$$g_k^{\mathrm{T}} d_k \leqslant -(1-\rho)\|g_k\|^2.$$

The above lemma shows that search direction sequence $\{d_k\}$ satisfies sufficient descent condition (6). If $\{d_k\}$ is also bounded then $\{d_k\}$ is uniformly gradient related.

Lemma 2.3. *For all* $k \geqslant m$,

$$\|d_k\|^2 \leqslant \beta_k^{(k)}\|g_k\|^2 + \sum_{i=2}^{m} \beta_{k-i+1}^{(k)}\|d_{k-i+1}\|^2 \leqslant \max_{2\leqslant i\leqslant m}(\|g_k\|^2, \|d_{k-i+1}\|^2).$$

The above two lemmas are easily proved [28].

3. Global convergence

In this section we will prove their global convergence under Assumption (H_1) and (H_2).

Lemma 3.1. *If* (H_2) *holds, then there exists* $\eta > 0$ *such that*

$$f_k - f_{k+1} \geqslant \eta \frac{\|g_k\|^4}{\gamma_k}, \tag{11}$$

where

$$\gamma_k = \max_{1\leqslant i\leqslant m}(\|g_k\|^2, \|d_{k-i+1}\|^2).$$

Proof. (i) For Armijo's line search rule (c), let

$$K_1 = \{k \| \alpha_k = s\}, \quad K_2 = \{k \| \alpha_k < s\},$$

we have

$$\begin{aligned} f_k - f_{k+1} &\geqslant s\mu_2 g_k^{\mathrm{T}} d_k \\ &= s\mu_2(1-\rho)\|g_k\|^2 \\ &\geqslant s\mu_2(1-\rho)\frac{\|g_k\|^4}{\gamma_k}, \quad k \in K_1 \end{aligned}$$

because of $\|g_k\|^2 \leqslant \gamma_k$.

For $k \in K_2$, $2\alpha_k \leqslant s$, by Armijo's line search rule, we have

$$f_k - f(x_k + 2\alpha_k d_k) < -2\mu_2\alpha_k g_k^{\mathrm{T}} d_k, \quad k \in K_2.$$

Using mean value theorem on the left-hand side of the above inequality, there exists $\theta_k \in [0,1]$ such that

$$-2\alpha_k g(x_k + 2\theta_k\alpha_k d_k)^{\mathrm{T}} d_k = f_k - f(x_k + 2\alpha_k d_k) < -2\mu_2\alpha_k g_k^{\mathrm{T}} d_k, \quad k \in K_2.$$

Thus

$$g(x_k + 2\alpha_k d_k)^{\mathrm{T}} d_k > \mu_2 g_k^{\mathrm{T}} d_k, \quad k \in K_2. \tag{12}$$

By (H_2), Cauchy–Schwartz inequality and (12), we obtain that

$$\begin{aligned} 2\alpha_k L\|d_k\|^2 &\geqslant \|g(x_k + 2\alpha_k d_k) - g_k\| \cdot \|d_k\| \\ &\geqslant (g(x_k + 2\alpha_k d_k) - g_k)^{\mathrm{T}} d_k \\ &> -(1-\mu_2) g_k^{\mathrm{T}} d_k, \quad k \in K_2. \end{aligned}$$

Therefore

$$\alpha_k > -\frac{(1-\mu_2)}{2L} \frac{g_k^{\mathrm{T}} d_k}{\|d_k\|^2}, \quad k \in K_2. \tag{13}$$

Also by Armijo's line search rule with Lemmas 2.2 and 2.3, we have

$$\begin{aligned} f_k - f_{k+1} &\geqslant -\mu_2 \alpha_k g_k^{\mathrm{T}} d_k \\ &> \frac{\mu_2(1-\mu_2)}{2L} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} \\ &\geqslant \frac{\mu_2(1-\mu_2)(1-\rho)^2}{2L} \frac{\|g_k\|^4}{\gamma_k}, \quad k \in K_2. \end{aligned}$$

Let

$$\eta = \min\left(s\mu_2(1-\rho), \frac{\mu_2(1-\mu_2)(1-\rho)^2}{2L}\right),$$

then (11) is proved.

(ii) For exact line search rule (a) and limited exact line search rule (b), let α_k^* be the exact stepsize or limited exact stepsize (in this case $\alpha_k^* \in [0, s]$), and α_k' be the stepsize obtained by Armijo's line search rule. From the above proof, we have

$$\begin{aligned} f_k - f_{k+1} &= f_k - f(x_k + \alpha_k^* d_k) \\ &\geqslant f_k - f(x_k + \alpha_k' d_k) \\ &\geqslant \eta \frac{\|g_k\|^4}{\gamma_k}. \end{aligned}$$

Eq. (11) is also proved.

(iii) For Goldstein's line search rule(d), By left-hand side inequality of Goldstein's line search rule, and the mean value theorem, there exists $\theta_k \in [0, 1]$ such that

$$\alpha_k g(x_k + \theta_k \alpha_k d_k)^{\mathrm{T}} d_k = f_{k+1} - f_k \geqslant \alpha_k \mu_2 g_k^{\mathrm{T}} d_k.$$

By (H_2) and Cauchy–Schwarz inequality, we have

$$\begin{aligned} \alpha_k L\|d_k\|^2 &\geqslant \|g(x_k + \alpha_k \theta_k d_k) - g_k\| \cdot \|d_k\| \\ &\geqslant [g(x_k + \alpha_k \theta_k d_k) - g_k]^{\mathrm{T}} d_k \\ &= -(1-\mu_2) g_k^{\mathrm{T}} d_k, \end{aligned}$$

i.e.,

$$\alpha_k \geqslant \frac{-(1-\mu_2)g_k^{\mathrm{T}}d_k}{L\|d_k\|^2}.$$

By the right-hand side inequality of Goldstein's line search rule, we have

$$\begin{aligned} f_k - f_{k+1} &\geqslant -\alpha_k \mu_1 g_k^{\mathrm{T}} d_k \\ &\geqslant \frac{\mu_1(1-\mu_2)}{L}\left(\frac{-g_k^{\mathrm{T}}d_k}{\|d_k\|}\right)^2 \\ &\geqslant \frac{\mu_1(1-\mu_2)}{L}\frac{(g_k^{\mathrm{T}}d_k)^2}{\gamma_k}\{\text{by Lemma 2.3}\} \\ &\geqslant \frac{\mu_1(1-\mu_2)}{L}\frac{(1-\rho)^2\|g_k\|^4}{\gamma_k}\ \{\text{by Lemma 2.2}\} \\ &\geqslant \eta\frac{\|g_k\|^4}{\gamma_k}, \end{aligned}$$

where

$$\eta = \frac{\mu_1(1-\mu_2)(1-\rho)^2}{L},$$

and (11) is proved.

(iv) For Wolfe's line search rule (e) (or strong Wolfe's line search rule), since $g_{k+1}^{\mathrm{T}}d_k \geqslant \mu_2 g_k^{\mathrm{T}}d_k$, by Cauchy–Schwartz inequality, we obtain

$$L\alpha_k\|d_k\| \geqslant \|g_{k+1}-g_k\|\cdot\|d_k\| \geqslant (g_{k+1}-g_k)^{\mathrm{T}}d_k \geqslant -(1-\mu_2)g_k^{\mathrm{T}}d_k,$$

thus

$$\alpha_k \geqslant -\frac{(1-\mu_2)}{L}\frac{g_k^{\mathrm{T}}d_k}{\|d_k\|^2}.$$

By the first inequality of Wolfe's line search rule, we have

$$\begin{aligned} f_k - f_{k+1} &\geqslant -\alpha_k \mu_1 g_k^{\mathrm{T}} d_k \\ &\geqslant \frac{\mu_1(1-\mu_2)}{L}\left(\frac{-g_k^{\mathrm{T}}d_k}{\|d_k\|}\right)^2 \\ &\geqslant \frac{\mu_1(1-\mu_2)}{L}\frac{(g_k^{\mathrm{T}}d_k)^2}{\gamma_k}\{\text{by Lemma 2.3}\} \\ &\geqslant \frac{\mu_1(1-\mu_2)}{L}\frac{(1-\rho)^2\|g_k\|^4}{\gamma_k}\ \{\text{by Lemma 2.2}\} \\ &\geqslant \eta\frac{\|g_k\|^4}{\gamma_k}, \end{aligned}$$

where

$$\eta = \frac{\mu_1(1-\mu_2)(1-\rho)^2}{L},$$

and (11) is proved. □

Theorem 3.1. *If* (H_1) and (H_2) *hold and Algorithm*(A) *generates an infinite sequence* $\{x_k\}$, *then*

$$\sum_{k=m}^{\infty} \frac{\|g_k\|^4}{\gamma_k} < +\infty, \tag{14}$$

where

$$\gamma_k = \max_{2 \leqslant i \leqslant m} (\|g_k\|^2, \|d_{k-i+1}\|^2).$$

Proof. Since $\{f_k\}$ is a decreasing sequence and has lower bound on the level set L_0, it is a convergent sequence. Therefore, Lemma 3.1 shows that (14) holds. The proof is complete. □

Lemma 3.2. *Suppose that* $\{x_k\}$ *generated by Algorithm* (A) *is bounded. Then*, $\{\|d_k\|^2\}$ *is also bounded.*

Proof. In fact, if $\{x_k\}$ has a bound, then $\{\|g_k\|^2\}$ has also a bound, M say. By the proof of Lemma 2.3, if $k \leqslant m-1$ then

$$\|d_k\|^2 = \|g_k\|^2 \leqslant M,$$

if $k = m$ then

$$\|d_k\|^2 \leqslant \max_{2 \leqslant i \leqslant m} \{\|g_k\|^2, \|d_{k-i+1}\|^2\} \leqslant M,$$

if $k > m$ then by induction process, we have

$$\|d_k\|^2 \leqslant \max_{2 \leqslant i \leqslant m} \{\|g_k\|^2, \|d_{k-i+1}\|^2\} \leqslant M.$$

Thus,

$$\|d_k\|^2 \leqslant M, \ \forall k \geqslant 1. \tag{15}$$

The proof is complete. □

Theorem 3.2. *The conditions in Theorem* 3.1 *hold*, *then*, *either*

$$\lim_{k \to \infty} \|g_k\| = 0$$

or $\{x_k\}$ *has no bound.*

Proof. If

$$\lim_{k \to \infty} \|g_k\| = 0$$

does not hold, then there exists an infinite subset $K_0 \subset \{m, m+1, \ldots\}$, $\varepsilon > 0$ *such that*

$$\|g_k\| > \varepsilon, \ \ k \in K_0,$$

thus

$$\frac{\varepsilon^4}{\gamma_k} \leqslant \frac{\|g_k\|^4}{\gamma_k}, \quad \forall k \in K_0.$$

By Theorem 3.1, we have

$$\sum_{k \in K_0} \frac{\varepsilon^4}{\gamma_k} \leqslant \sum_{k=m}^{+\infty} \frac{\|g_k\|^4}{\gamma_k} < +\infty.$$

Then, there exists at least one $i_0 : 2 \leqslant i_0 \leqslant m$ such that

$$\lim_{k \in K_0, k \to \infty} \|d_{k-i_0+1}\| = +\infty.$$

Lemma 2.3 shows that $\{\|g_k\|\}$ has no bound. By Assumption (H_2), we have

$$L\|x_k - x_0\| \geqslant \|g_k - g_0\|,$$

thus $\{x_k\}$ has no bound.

Conversely, if $\{x_k\}$ has a bound, then $\{\|g_k\|\}$ has also a bound, by Lemma 3.2, there exists an $M > 0$ such that $\gamma_k \leqslant M$. By Theorem 3.1,

$$\sum_{k=m}^{\infty} \frac{\|g_k\|^4}{M} \leqslant \sum_{k=m}^{\infty} \frac{\|g_k\|^4}{\gamma_k} < +\infty.$$

Thus $\lim_{k\to\infty} \|g_k\|^4 = 0$, i.e., $\lim_{k\to\infty} \|g_k\| = 0$. This ends the proof. □

Corollary 3.1. *If the level set* L_0 *of objective function* $f(x)$ *is bounded and Assumption* (H_2) *holds, then*

$$\lim_{k\to\infty} \|g_k\| = 0.$$

Note: Although the unboundedness of $\{\|g_k\|\}$ implies the unboundedness of $\{x_k\}$, the unboundedness of $\{x_k\}$ can not imply the unboundedness of $\{\|g_k\|\}$. Indeed, we can prove that whether $\{x_k\}$ is bounded or not, $\{\|g_k\|\}$ is always bounded.

Lemma 3.3. *For* $k \geqslant 1$,

$$\|d_k\|^2 \leqslant \max_{1 \leqslant i \leqslant k} \{\|g_i\|^2\}.$$

Proof. If $k \leqslant m$, then the conclusion is obvious. If $k > m$, then by induction process, we obtain the conclusion. The proof is complete. □

In fact, we can prove the following result.

Lemma 3.4. *If the conditions in Theorem* 3.1 *hold, then both* $\{\|g_k\|\}$ *and* $\{\|d_k\|\}$ *has a bound.*

Proof. By Lemmas 2.3 and 3.3, it is sufficient to prove that $\{\|g_k\|\}$ has a bound. Using inverse proof, if there exists an infinite sequence K' satisfying

$$\|g_k\|^2 \to +\infty (k \in K',\ k \to +\infty),$$

then, there exists an infinite sequence $K'' \subseteq K'$ such that

$$\|g_k\|^2 \geqslant \max_{2 \leqslant i \leqslant m} \{\|g_k\|^2, \|d_{k-i+1}\|^2\} = \gamma_k,\ \ \forall k \in K'',$$

where γ_k is as in (11). By Lemma 3.3 and Theorem 3.1, we can obtain

$$\begin{aligned}\sum_{k \in K''} \|g_k\|^2 &\leqslant \sum_{k \in K''} \frac{\|g_k\|^4}{\gamma_k} \\ &\leqslant \sum_{k=m}^{\infty} \frac{\|g_k\|^4}{\gamma_k} < +\infty.\end{aligned}$$

This is a contradiction. Thus $\{\|g_k\|^2\}$ has a bound. □

Lemma 3.4 shows that whether $\{x_k\}$ has a bound or not, $\{\|g_k\|\}$ and $\{\|d_k\|\}$ are bounded. The following theorem is the main theorem of this section.

Theorem 3.3. *If the conditions in Theorem* 3.1 *hold, then*

$$\lim_{k \to +\infty} \|g_k\| = 0. \tag{16}$$

Proof. Let

$$\delta = \limsup_{k \to +\infty} \|g_k\|^2.$$

By Lemma 3.4, we have $\delta < +\infty$. It suffices to prove $\delta = 0$. If

$$\delta > 0, \tag{17}$$

then, by Lemma 3.3, there exists k' and an infinite subset, K' say, such that

$$\|g_k\|^2 \geqslant \tfrac{1}{2}\delta, \quad \gamma_k \leqslant M, \quad \forall k \geqslant k',\ k \in K'.$$

By Theorem 3.1, we have

$$\begin{aligned}\sum_{k \in K', k \geqslant k'} \frac{\frac{1}{4}\delta^2}{M} &\leqslant \sum_{k \in K', k \geqslant k'} \frac{\|g_k\|^4}{M} \\ &\leqslant \sum_{k \in K', k \geqslant k'} \frac{\|g_k\|^4}{\gamma_k}\end{aligned}$$

$$\leqslant \sum_{k=m}^{\infty} \frac{\|g_k\|^4}{\gamma_k}$$
$$< +\infty.$$

This is a contradiction with (17), thus $\delta = 0$. The proof is complete. □

In fact, if we can prove that $\{d_k\}$ has a bound, then $\{d_k\}$ is uniformly gradient-related to $\{x_k\}$, by Lemma 2.1, Algorithm (A) has global convergence. As a result, the global convergence can also be proved by using Lemmas 2.1 and 3.4.

4. Further convergence results

As we can see that (H_2) implies (H_3), we can obtain global convergence if (H_1) and (H_3) hold. As a result, the convergence results in Section 3 are apparent.

Lemma 4.1. *If* (H_1) *and* (H_3) *hold, Algorithm* (A) *with inexact line search* (c), (d), *or* (e) *generates an infinite sequence* $\{x_k\}$, *then* $\{\|g_k\|\}$ *and* $\{\|d_k\|\}$ *are bounded.*

Proof. By Lemma 3.3, it is sufficient to prove that $\{\|g_k\|\}$ is bounded. Let

$$\delta_k = \max_{1 \leqslant j \leqslant k} \{\|g_j\|\}. \tag{18}$$

If $\{\|g_k\|\}$ has no bound, then

$$\lim_{k \to +\infty} \delta_k = +\infty,$$

there must exist an infinite subset $N \subseteq \{2, 3, \ldots\}$ such that $\delta_k = \|g_k\|$, $\forall k \in N$ and

$$\|g_k\| = \delta_k \to +\infty (k \in N, k \to +\infty). \tag{19}$$

By (H_1), Lemma 2.2 and inexact line search rule (c)–(e), we have

$$\sum_{k=m}^{+\infty} \alpha_k \|g_k\|^2 < +\infty, \tag{20}$$

thus

$$\sum_{k \in N} \alpha_k \|g_k\|^2 \leqslant \sum_{k=m}^{+\infty} \alpha_k \|g_k\|^2 < +\infty. \tag{21}$$

By Lemma 2.3, we obtain

$$\|d_k\|^2 \leqslant \delta_k^2 = \|g_k\|^2, \ \forall k \in N. \tag{22}$$

Using (21) and (22), we have

$$\sum_{k \in N} \alpha_k \|d_k\|^2 < +\infty, \tag{23}$$

thus

$$\alpha_k\|d_k\|^2 \to 0 \quad (k \in N, k \to +\infty). \tag{24}$$

By Lemma 2.2 and Cauchy–Schwarz inequality, we obtain

$$\|g_k\| \cdot \|d_k\| \geqslant -g_k^{\mathrm{T}} d_k \geqslant (1-\rho)\|g_k\|^2,$$

hence

$$\|d_k\| \geqslant (1-\rho)\|g_k\|. \tag{25}$$

By (19) and (25) one has

$$\|d_k\| \to +\infty \quad (k \in N, k \to +\infty),$$

and noting (24), we have

$$\alpha_k\|d_k\| \to 0 \quad (k \in N, k \to +\infty). \tag{26}$$

(i) For Armijo line search rule (c), let $N_1 = \{k|\alpha_k \geqslant s\}$, $N_2 = \{k|\alpha_k < s\}$, then

$$s\sum_{k\in N_1} \alpha_k\|g_k\|^2 + \sum_{k\in N_2} \alpha_k\|g_k\|^2 < +\infty.$$

If N_1 is an infinite subset then $\|g_k\| \to 0$ ($k \in N_1, k \to +\infty$), thus we can take $N \subseteq N_2$, therefore $\forall k \in N$, we have $2\alpha_k \leqslant s$, Armijo's line search rule implies that

$$f_k - f(x_k + 2\alpha_k d_k) < -2\mu_2\alpha_k g_k^{\mathrm{T}} d_k, \ \forall k \in N.$$

Using mean value theorem, there exists $\theta_k \in [0,1]$ such that

$$-2\alpha_k g(x_k + 2\alpha_k\theta_k d_k)^{\mathrm{T}} d_k = f_k - f(x_k + 2\alpha_k d_k) < -2\mu_2\alpha_k g_k^{\mathrm{T}} d_k, \quad \forall k \in N,$$

thus

$$g(x_k + 2\alpha_k\theta_k d_k)^{\mathrm{T}} d_k > \mu_2 g_k^{\mathrm{T}} d_k. \tag{27}$$

From Lemma 2.2, (27), and (22), we have

$$\begin{aligned}(1-\rho)(1-\mu_2)\|g_k\|^2 &\leqslant (1-\mu_2)(-g_k^{\mathrm{T}} d_k)\\ &\leqslant (g(x_k + 2\alpha_k\theta_k d_k) - g_k)^{\mathrm{T}} d_k\\ &\leqslant \|d_k\| \cdot \|g(x_k + 2\alpha_k\theta_k d_k) - g_k\|\\ &\leqslant \|g_k\| \cdot \|g(x_k + 2\alpha_k\theta_k d_k) - g_k\|, \quad \forall k \in N\end{aligned}$$

thus

$$(1-\rho)(1-\mu_2)\|g_k\| \leqslant \|g(x_k + 2\alpha_k\theta_k d_k) - g_k\|, \ \forall k \in N.$$

By (H_3), (26) and the above inequality, we have $\|g_k\| \to 0$ ($k \in N, k \to +\infty$), which contradicts to (19). This shows that $\{\|g_k\|\}$ is bounded. $\{\|g_k\|\}$ has a bound, so does $\{\|d_k\|\}$ from Lemma 2.3.

(ii) For Goldstein line search rule: by the proof of Lemma 3.1, we have

$$g(x_k + \alpha_k\theta_k d_k)^{\mathrm{T}} d_k \geqslant \mu_2 g_k^{\mathrm{T}} d_k, \tag{28}$$

where $\theta_k \in [0, 1]$, thus

$$\begin{aligned}(1-\rho)(1-\mu_2)\|g_k\|^2 &\leqslant (1-\mu_2)(-g_k^{\mathrm{T}} d_k)\\ &\leqslant (g(x_k+\alpha_k\theta_k d_k)-g_k)^{\mathrm{T}} d_k\\ &\leqslant \|d_k\|\cdot\|g(x_k+\alpha_k\theta_k d_k)-g_k\|\\ &\leqslant \|g_k\|\cdot\|g(x_k+\alpha_k\theta_k d_k)-g_k\|, \ \forall k\in N\end{aligned}$$

therefore,

$$(1-\rho)(1-\mu_2)\|g_k\| \leqslant \|g(x_k+\alpha_k\theta_k d_k)-g_k\|.$$

By (H_3), (26) and the above inequality, we have $\|g_k\| \to 0 (k \in N, k \to +\infty)$, which contradicts to (19). This shows that $\{\|g_k\|\}$ is bounded.

(iii) For strong Wolfe line search rule: since

$$g(x_k+\alpha_k d_k)^{\mathrm{T}} d_k \geqslant \mu_2 g_k^{\mathrm{T}} d_k, \tag{29}$$

we have

$$\begin{aligned}(1-\rho)(1-\mu_2)\|g_k\|^2 &\leqslant (1-\mu_2)(-g_k^{\mathrm{T}} d_k)\\ &\leqslant (g(x_k+\alpha_k d_k)-g_k)^{\mathrm{T}} d_k\\ &\leqslant \|d_k\|\cdot\|g(x_k+\alpha_k d_k)-g_k\|\\ &\leqslant \|g_k\|\cdot\|g(x_k+\alpha_k d_k)-g_k\|, \ \forall k\in N\end{aligned}$$

thus

$$(1-\rho)(1-\mu_2)\|g_k\| \leqslant \|g(x_k+\alpha_k d_k)-g_k\|.$$

By (H_3), (23) and the above inequality, we have $\|g_k\| \to 0 (k \in N, k \to +\infty)$, which contradicts to (19). This shows that $\{\|g_k\|\}$ has a bound. □

Theorem 4.1. *If* (H_1) *and* (H_3) *hold, and Algorithm* (A) *with inexact line search rule* (c)–(e), *generates an infinite sequence* $\{x_k\}$, *then*

$$\lim_{k\to+\infty} \|g_k\| = 0. \tag{30}$$

Proof. It is easy to prove by using Lemma 4.1. In fact, Lemma 4.1 guarantees that $\{\|d_k\|\}$ has a bound, say M. By (27)–(29), and Lemma 2.2, we have

$$\begin{aligned}M\|g(x_k+y_k)-x_k\| &\geqslant \|g(x_k+y_k)-g_k\|\cdot\|d_k\|\\ &\geqslant (g(x_k+y_k)-g_k)^{\mathrm{T}} d_k\\ &\geqslant -(1-\mu_2) g_k^{\mathrm{T}} d_k\\ &\geqslant \mu_2(1-\rho)\|g_k\|^2,\end{aligned}$$

where $y_k = 2\alpha_k\theta_k d_k$ for (27) with $\theta_k \in [0,1]$, $y_k = \alpha_k\theta_k d_k$ for (28) with $\theta_k \in [0,1]$, $y_k = \alpha_k d_k$ for (29), respectively. Therefore

$$M\|g(x_k + y_k) - g_k\| \geqslant \mu_2(1-\rho)\|g_k\|^2. \tag{31}$$

For inexact line search rule (c)–(e), we have

$$\sum_{k=m}^{\infty} \alpha_k \|g_k\|^2 < +\infty. \tag{32}$$

This implies that

$$\alpha_k\|g_k\|^2 \to 0 \quad (k \to \infty). \tag{33}$$

If (30) does not hold, then there must exist an infinite subset $K \subseteq \{m, m+1, m+2, \ldots\}$ and an $\varepsilon > 0$ such that

$$\|g_k\| > \varepsilon, \ \ k \in K. \tag{34}$$

From (33) and (34) we have

$$\alpha_k \to 0 \quad (k \in K, \ k \to +\infty). \tag{35}$$

Lemma 4.1 shows that $\{\|d_k\|\}$ has a bound, by (35) we obtain

$$\alpha_k d_k \to 0 \quad (k \in K, \ k \to +\infty). \tag{36}$$

Thus

$$\|y_k\| \leqslant \alpha_k\|d_k\| \to 0 \quad (k \in K, \ k \to +\infty). \tag{37}$$

By Assumption (H_3), (31) and (37) we have

$$\|g_k\|^2 \leqslant \|g(x_k + y_k) - g_k\| \cdot \frac{M}{\mu_2(1-\rho)} \to 0 \quad (k \in K, \ k \to \infty).$$

This is a contradiction with (34). The contradiction shows that (30) holds. □

Note: There is other simple proof. From Lemma 4.1, search direction sequence $\{\|d_k\|\}$ is uniformly gradient-related to $\{x_k\}$. By Lemma 2.1, any limit point x^* of $\{x_k\}$ is a critical point of (UP).

Corollary 4.1. *If Assumptions* (H_1) *and* (H_2) *hold, Algorithm* (A) *with inexact line search rule generates an infinite sequence* $\{x_k\}$, *then* (30) *holds.*

Proof. It is easy to prove because Assumption (H_2) implies Assumption (H_3).

Remark. Obviously, in this section we show that the new algorithm has convergence property under weak conditions. Specifically, we do not need the boundedness of level set to guarantee the convergence, i.e., the sequence $\{x_k\}$ generated by new algorithm may be an unbounded sequence, however

$$\lim_{k\to+\infty} \|g_k\| = 0$$

always holds. For instance, the new algorithm may solve the following minimization problem:

$$\min f(x) = \mathrm{e}^{-x}, \quad x \in R.$$

Moreover, we need only the uniformly continuous of objective function f to prove the global convergence of new algorithm.

For further research we should study the global convergence in the case of exact line searches under Assumption (H_1) and (H_3).

5. Linear convergence rate

Assumption. (H_4): f is uniformly convex and twice continuously differentiable.

In fact, Assumption (H_4) implies (H_1) and (H_2) and thus implies (H_3).

Lemma 5.1. *If* (H_4) *holds, then f has the following properties*:

(1) *f has a unique minimizer on $\mathbb{R}^n$, say x^*.*
(2) *The level set $L_0 = \{x \mid f(x) \leqslant f(x_1)\}$ is bounded.*
(3) *There exist $m' > 0$, $\quad M' > 0$ such that*

$$m'\|x - x^*\|^2 \leqslant f(x) - f(x^*) \leqslant \tfrac{1}{2} M'\|x - x^*\|^2,$$

$$m'\|x - x^*\| \leqslant \|g(x)\| \leqslant M'\|x - x^*\|.$$

(4) *Assumptions* $(H_1), (H_2)$ *and* (H_3) *hold.*

Proof. The proof follows from mean value theorem, and here is omitted.

Theorem 5.1. *If* (H_4) *holds, then $\{x_k\} \to x^*$, where $x*$ is the unique minimizer of f. Further, either there exists an infinite subset $K \subset \{m, m+1, \ldots\}$ and $i_0 : 2 \leqslant i_0 \leqslant m$ such that*

$$\lim_{k\to\infty, k\in K} \frac{\|g_k\|}{\|d_{k-i_0+1}\|} = 0$$

or

$$R_1\{x_k\} = \limsup_{k\to\infty} \|x_k - x^*\|^{\frac{1}{k}} < 1.$$

Proof. By Lemma 5.1 and Theorem 3.3, we have

$$\lim_{k\to\infty} x_k = x^*.$$

By Lemma 3.1, if $\{\|d_{k-i+1}\|/\|g_k\|\}$ has no bound for at least one $i : 2 \leqslant i \leqslant m$, there exists an infinite subset K and i_0 such that

$$\lim_{k\to\infty, k\in K} \frac{\|d_{k-i_0+1}\|}{\|g_k\|} = \infty,$$

thus

$$\lim_{k\to\infty, k\in K} \frac{\|g_k\|}{\|d_{k-i_0+1}\|} = 0.$$

If $\{\|d_{k-i+1}\|/\|g_k\|\}$ has a bound for all $2 \leqslant i \leqslant m$ and all $k \geqslant m$, i.e., there exists $\mu_0 > 0$ such that

$$\frac{\|d_{k-i+1}\|}{\|g_k\|} \leqslant \mu_0, \ \forall i: 2 \leqslant i \leqslant m,$$

hence

$$\frac{\|d_{k-i+1}\|^2}{\|g_k\|^2} \leqslant \mu_0^2, \quad \forall i: 2 \leqslant i \leqslant m.$$

Therefore

$$f_k - f_{k+1} \geqslant \eta_0 \|g_k\|^2, \tag{38}$$

where

$$\eta_0 = \frac{\eta}{\max(1, \mu_0^2)},$$

and η is, respectively, taken as in the proof of Lemma 3.1.

By (38), the remainder proof follows from [2]. □

6. Numerical experiments

For nonquadratic objective function in (UP), we use inexact line search rules (c), (d), or (e) to choose the step-size α_k in steepest descent method, FR, PRP, HS, CD, DY methods, etc. The new method in the paper is denoted by NM, the steepest descent method by SM. All these methods have the same property that avoids the overhead and evaluation of second derivative of f, the storage and computation of matrix associated with Newton-type methods.

$$|f_k - f^*| \leqslant \text{eps}$$

in which eps is the computational precision.

We choose the following problems for our numerical experiments.

Test 1 (Huang and Chambliss [17]).

$$f(x) = (x_1 + 10x_2)^4 + 5(x_3 - x_4)^4 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4,$$

$$x_0 = (2, 2, -2, -2)^{\mathrm{T}}, \ x^* = (0, 0, 0, 0)^{\mathrm{T}}, \ f^* = 0.$$

Test 2 (Dixon [8]).

$$f = (1 - x_1)^2 + (1 - x_{10})^2 + \sum_{i=1}^{9} (x_i^2 - x_{i+1})^2,$$

$$x_0 = (-2, \ldots, -2)^{\mathrm{T}}, \ x^* = (1, \ldots, 1)^{\mathrm{T}}, \ f^* = 0.$$

Table 1
The iterative number for attaining the same precision

P	NM	SM	FR	PRP	HS	CD	DY
1	16/16	67/65	14/15	17/13	17/18	15/21	17/20
2	34/30	146/104	F/167	75/98	F/56	72/91	45/42
3	44/22	F/F	58/77	63/65	64/64	64/64	F/F
4	344/352	498/421	F/F	392/336	476/409	352/363	372/428
5	5/5	8/9	5/7	7/7	6/8	9/7	3/5
6	21/18	32/29	47/36	23/20	28/25	26/23	25/23
7	18/18	56/50	28//31	32/29	38/32	45/33	71/42
8	33/23	47/43	55/50	62/55	71/62	35/38	48/48
9	35/45	97/72	35/47	54/53	63/55	65/38	67/41
10	67/60	56/62	73/64	56/50	74/55	63/77	87/66
T	123/116	345/318	217/201	463/424	342/398	321/301	317/297

Test 3. Extended Powell function [13]:

$$f = \sum_{i=1}^{n-3} [(x_i + 10x_{i+1})^2 + 5(x_{i+2} - x_{i+3})^2 + (x_{i+1} - 2x_{i+2})^4 + 10(x_i - x_{i+3})^4],$$

$$x_0 = (3, -1, 0, 1, \ldots, 3, -1, 0, 1)^{\mathrm{T}}, \;\; x^* = (0, 0, \ldots, 0)^{\mathrm{T}}, \;\; f^* = 0.$$

Test 4. Take $n = 10,000$ in Test 3.

Tests 5–10 are, respectively, the Problems 25, 26, 28, 29, 30, 33 from Appendix A of Himmelblau's book [18].

Large computation and comparison show that taking $m = 3$, $\mu_1 = 0.35$, $\mu_2 = 0.75$, $\rho = 0.95$ and $\beta_{k-i+1} = s_k^i$, $i = 2, \ldots, m$, $\text{eps} = 10^{-7}$ is available for many practical problems. Armijo line search rule is always used in the algorithm in the paper. In Armijo's line search rule, we take $s = 2\|g(x_0)\|$. Other line search rules may also be used, we do not list the results of the implementation of algorithms with line search rules (d) and (e). The numerical implementation results of algorithm with line search rule (c) are reported in Table 1.

In the above table, 'F' means the corresponding methods fail in the case. 'P' denotes the test problems. 'T' denotes the total CPU time (seconds) of solving all the 10 problems. The second number in the above table denotes the results of algorithm using generalized Armijo's line search rule with $\beta = 0.75$.

We can see that the new algorithm needs fewer iterations when reaching the same precision. We can also find that the new algorithm converges more stably than other methods in many situations. Moreover, the implementation of the new algorithm with generalized Armijo's line search seems to have different results. We use generalized Armijo's line search with $\beta = 0.38,\ 0.55,\ 0.75$, respectively. It can be find that all algorithms seem to perform more stably and effectively in the case of using generalized Armijo's line search with $\beta = 0.75$ and $s = \|g(x_0)\|$. However, the new algorithm in the paper is more efficient and stable than other similar methods (Table 2).

Table 2
The function evaluation number for attaining the same precision

P	NM	SM	FR	PRP	HS	CD	DY
1	227/225	2431/1652	820/720	743/679	932/1024	436/554	312/412
2	796/612	609/579	F/679	595/520	F/426	652/552	735/814
3	145/123	F/F	663/732	934/559	715/682	367/327	F/F
4	2145/ 2005	3433/ 3172	F/ F	2981/ 2872	7485/ 7169	3100/ 2987	3836/ 3624
5	34/36	45/39	36/41	67/55	43/39	63/63	47/66
6	56/43	68/56	59/54	71/68	76/71	64/76	62/58
7	76/78	64/72	87/81	81/72	97/83	63/83	64/67
8	34/34	53/53	72/58	65/54	79/62	49/59	50/82
9	55/44	59/51	61/52	72/66	86/71	59/67	78/73
10	121/103	145/125	214/217	310/326	239/229	203/198	317/301

The computational results show that the new method in the paper is very efficient in practice. Firstly, like FR, PRP, HS, CD, DY, and Steepest descent method, the new method in the paper avoids the evaluation of second derivatives of f. Secondly, the storage of any matrix associated with Newton type method is avoided at each iteration. The last but not the least important thing is that the new method needs fewer iterations, fewer evaluations of f and g than FR, PRP, HS, CD, DY, and steepest descent method in many situations, etc., when the iterative process reaches the same precision. The new method uses less CPU time than other methods mentioned in the paper.

From the table we can see that other methods may fail in some cases, while the new algorithm always converges. It is obvious that the new algorithm uses less total CPU time than other methods do. Moreover numerical experiments also show that the new algorithm always converges stably. It seems to be suitable to solve ill-conditioned problems and suitable to solve large-scale minimization problems.

7. Conclusion

In this paper, a new gradient-related algorithm for solving large-scale unconstrained optimization problems is proposed. The new algorithm is a kind of line search method. The basic idea is to choose a combination of the current gradient and some previous search directions as a new search direction and to find a step-size by using various inexact line searches. Using more information at the current iterative step may improve the performance of the algorithm. This motivates us to find some new gradient algorithms which may be more effective than standard conjugate gradient methods. Uniformly gradient-related conception is useful and it can be used to analyze global convergence of the new algorithm. We proved the global convergence under weak conditions using another technique. Numerical experiments show that the new algorithm converge more stably and is more efficient and superior to other similar methods in many situations. The new algorithm is expected to solve ill-conditioned problems.

For further research, we should study the nonmonotone methods for unconstrained optimization (for example [13]) and compare the performance of these two classes of methods. Moreover, more numerical experiments for large practical problems still be done in the future. How to choose the parameters in the algorithm is another aspect of future investigation.

Acknowledgements

The authors wish to thank the anonymous referees and the editors for their constructive comments and valuable suggestions which improved the paper.

References

[1] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York, 1982.
[2] A.I. Cohen, Stepsize analysis for descent methods, J. Optim. Theory Appl. 33 (2) (1981) 187–205.
[3] J.W. Cantrell, On the relation between the memory gradient and the Fletcher–Reeves method, J. Optim. Theory Appl. 4 (1) (1969) 67–71.
[4] E.E. Cragg, A.V. Levy, Study on a supermemory gradient method for the minimization of functions, J. Optim. Theory Appl. 4 (3) (1969) 191–205.
[5] Y.H. Dai, Y. Yuan, Nonlinear Conjugate Gradient Method, Shanghai Science and Technology Press, Shanghai, 1999.
[6] Y.H. Dai, Y. Yuan, A nonlinear conjugate gradient with a strong Wolfe convergence property, SIAM J. Optim. 10 (1999) 177–182.
[7] Y.H. Dai, Y. Yuan, Convergence properties of the Fletcher–Reeves method, IMA J. Numer. Anal. 16 (1996) 155–164.
[8] L.C.W. Dixon, Conjugate directions without line searches, J. Institute of Math. Appl. 11 (1973) 317–328.
[9] R. Fletcher, Practical Methods of Optimization, Vol. 1, Unconstrained Optimization, New York, Wiley, 1987.
[10] R. Flecther, C.M. Revees, Function minimization by conjugate gradients, Comput. J. 7 (1964) 149–154.
[11] L. Grippo, S. Lucidi, A globally convergent version of the Polak–Ribière conjugate gradient method, Math. Programming 78 (1997) 375–391.
[12] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim. 2 (1) (1992) 21–42.
[13] L. Grippo, F. Lampariello, S. Lucidi, A class of nonmonotone stability methods in unconstrained optimization, Numer. Math. 62 (1991) 779–805.
[14] R. Horst, P.M. Pardalos, N.V. Thoai, Introduction to Global Optimization, Kluwer Academic Publishers, Dordrecht, Holland, 1995.
[15] Y.F. Hu, C. Storey, Global convergence restart for conjugate gradient methods, J. Optim. Theory Appl. 72 (1) (1991) 399–405.
[16] M.R. Hestenes, Conjugate Direction Methods in Optimization, Springer, New York Inc., 1980.
[17] H.Y. Huang, J.P. Chambliss, Quadratically convergent algorithms and one-dimensional search schemes, J. Optim. Theory Appl. 11 (2) (1973) 175–188.
[18] D.M. Himmelblau, Applied Nonlinear Programming, McGraw-Hill Book Company, New York, 1972.
[19] M.R. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bureau Stand. 29 (1952) 409–439.
[20] D.C. Luenberger, Linear and Nonlinear Programming, 2nd Edition, Addition-Wesley, Reading, MA, 1989.
[21] A. Miele, J.W. Cantrell, Study on a memory gradient method for the minimization of functions, J. Optim. Theory Appl. 3 (6) (1969) 459–470.
[22] O.L. Mangasarian, Nonlinear Programming, McGraw-Hill, New York, 1969.
[23] J. Nocedal, S.J. Wright, Numerical Optimization, New York, Springer, Berlin, 1999.
[24] E. Polak, Optimization: Algorithms and Consistent Approximations, Springer, New York, 1997.
[25] E. Polak, G. Ribière, Note sur la convergence des methodes de directions conjuguees, Revue Francaise d'Inform. Rec. Oper. 16 (1969) 35–43.

[26] B.T. Polyak, The conjugate gradient algorithm in extremal problems, U.S.S.R. Comput. Math. Math. Phys. 9 (1961) 94–112.
[27] D.F. Shanno, Conjugate gradient methods with inexact searches, Math Oper Res. 3 (1978) 244–256.
[28] Z.J. Shi, A new memory gradient method under exact line search, Asian Pacific J. Oper. Res. 20 (2) (2003).
[29] Z.J. Shi, Restricted PR conjugate gradient method and its global convergence, Adv. Math. 31 (1) (2002) 47–55.
[30] Z.J. Shi, A super-memory gradient method for unconstrained minimization, J. Eng. Math. 17 (2) (2000) 99–104.
[31] M.N. Vrahatis, G.S. Androulajis, J.N. Lambrinos, G.D. Magoulas, A class of gradient unconstrained minimization algorithms with adaptive stepsize, J. Comput. Appl. Math. 114 (2) (2000) 367–386.
[32] M.A. Wolfe, C. Viazminsky, Supermemory Descent Methods for Unconstrained Minimization, J. Optim. Theory Appl. 18 (4) (1976) 455–468.